What effects do errors in the classification and diagnosis of data have on the epidemiological studies in which they occur? This problem was discussed in the July issue of the Journal by Diamond and Lilienfeld. Their interpretation is disputed here. Dr. Lilienfeld will reply in the next issue of the Journal.

ERRORS IN THE INTERPRETATION OF ERRORS IN EPIDEMIOLOGY

David J. Newell, M.A., Ph.D.

IN THEIR paper, "Effects of Errors in Classification and Diagnosis in Various Types of Epidemiological Studies," Diamond and Lilienfeld[1] perpetuate a common error in the analysis of false-positives and false-negatives. Since in so doing they appear to disprove a well-known statistical result that misclassification tends to decrease true differences,[2] a note in refutation is called for.

In epidemiological studies,[3] a patient who has a condition on examination (i.e., a positive) may report that he does not have it. He is called a false-negative. Similarly, a patient who does not have the condition on examination (i.e., a negative) may report that he has it. He is called a false-positive. Notice that the population "at risk" of being false-negatives consists of persons who actually have the condition, and the population at risk of being false-positives are those who do not have it. Other definitions of false-positives and false-negatives lead to errors in interpretation, as will be shown later.

The importance of these definitions is not in the interpretation of the original investigation, in which both the true status (on examination) and the reported status of each individual is known, but in the interpretation of other studies in which only the reported status is known.

One of the most important situations in which the method is necessary is that in which some relatively inexpensive screening device (postal questionnaire, interview by nonmedical personnel, mass radiography) is used to estimate the prevalence of a condition whose true diagnosis is more expensive to establish.

Correct Method of Analysis

Suppose that the original investigation yields the results in Table 1. Of the n_1 true positives, b report themselves as negative. The false-negative rate is thus b/n_1, which we call β. Similarly, the false-positive rate c/n_2 is called α. (The symbols α and β are chosen by analogy with the notation for statistical errors of the first and second kind.)

Suppose that another investigation establishes only that in a total population of N, there are n_3 reported positives, and n_4 $(=N-n_3)$ reported negatives. Our task is to use the information on α and β from the first investi-

Table 1—Cases Classified by True Status and Reported Status in Original Investigation

		Reported Status		
		Positive	Negative	Total
True status on examination	positive	a	b	n_1
	negative	c	d	n_2

False-negative rate $= b/n_1 = \beta$.
False-positive rate $= c/n_2 = \alpha$.

gation to find the true number of cases in the second investigation.

This can be done by building up the missing information in Table 2 as follows:

1. Initially, the table contains only the data n_3, n_4, N.
2. Enter n_5 as the unknown number of true cases.
3. Hence β n_5 can be entered as the number of false-negatives.
4. Likewise, $\alpha(N-n_5)$ can be entered as the number of false-positives.
5. By subtraction, the number of "true positives" is $n_3-\alpha(N-n_5)$.
6. Now adding across the top row, we find $n_3-\alpha(N-n_5)+\beta n_5=n_5$.

Therefore the unknown number of true cases, n_5, is given by

$$n_5=\frac{n_3-\alpha N}{1-\alpha-\beta}. \qquad (1)$$

Notice that the reported prevalence rate

$$n_3/N=\alpha+(1-\alpha-\beta)n_5/N \qquad (2)$$

must always lie somewhere between α and $1-\beta$.

If the true prevalence is 100 per cent, a proportion β of these cases will report themselves as negative, giving a reported prevalence of $1-\beta$, while if the true prevalence is 0 per cent, a proportion α of the population will report themselves as positive, giving a reported prevalence of α. For intermediate values of true prevalence between 0 and 100 per cent, the corresponding reported prevalence figures will lie proportionately between these extremes of α and $1-\beta$.

A Common Mistaken View of Errors

Diamond and Lilienfeld[1] have not used the false-negative and false-positive rates defined above, but instead have transferred rates based on the reported status. For example, they use the ratio $a/(a+c)$ in the notation of Table 1 to define $P(T+|S+)$, the probability that a person who says he has the characteristic actually has it. Now this probability is not independent of the true prevalence, and is thus not applicable to another investigation where the true prevalence rate may be different. This is most clearly seen by supposing that the original investigation had been carried out in an area where all cases were truly positive. Then c and d must be zero, and $a/(a+c)=1$. If this figure is transferred to another investigation, it would imply that there are no errors in classification of those reporting themselves as positive.

Table 2—Derived Classification of Cases in Second Investigation, Given Only the Reported Status Numbers n_3 and n_4

		Reported Status		
		Positive	Negative	Total
True status on examination	positive	$n_3-\alpha(N-n_5)$	βn_5	n_5
	negative	$\alpha(N-n_5)$		$N-n_5$
	Total	n_3	n_4	N

Reanalysis of the Circumcision Data

From Lilienfeld and Graham's original investigation,[4] we find that of the 84 actually circumcised patients, 47 stated that they were not circumcised, thus giving a false-negative rate $\beta=$ 56.0 per cent. Of the 108 who were not circumcised, 19 claimed that they were, giving a false-positive rate of $\alpha=17.6$ per cent.

Diamond and Lilienfeld attempt to use these data to correct a study of Wynder's,[5] in which cases and controls had stated percentages circumcised of 5 per cent and 14 per cent, respectively. Now neither of these figures is in the range $\alpha=17.6$ per cent to $1-\beta=44.0$ per cent. Hence, Wynder's data are inconsistent with Lilienfeld and Graham's. Even if the true proportion circumcised in Wynder's study had been 0 per cent, then, apart from sampling variations, 17.6 per cent of them would have reported themselves circumcised had the data from Lilienfeld and Graham been applicable. Without precise knowledge of the way in which the reported data were obtained in the two surveys, we cannot be certain how the discrepancy arose, but Lilienfeld and Graham clearly indicate that there were certain differences in the groups studied and the reporting technics. In particular, Wynder's data were provided by the spouses, who might have a lower false-positive rate.

As an added refutation of Diamond and Lilienfeld's argument, it can be pointed out that if their second method of deriving stated percentages is applied to the derived true percentages obtained from their first method, it does not lead back to the original stated percentages.

Comparison Between Two Proportions with Misclassification

We can now reinstate the proposition that misclassification always tends to reduce the apparent difference between two proportions, on which Diamond and Lilienfeld have thrown doubt.

If the probabilities α and β of misclassification are the same for each of two groups (e.g., cases and controls) but the true prevalence rates p_1 and p_2 are different, then the reported prevalence rates can be expected to differ by less than p_1-p_2.

From Equation (2) the expected value of the difference between two reported rates is $(1-\alpha-\beta)$ times the difference between the true rates. This factor is

Table 3—Percentage Distribution of 1,776 Males and 2,064 Females by Chronic Bronchitis Status, as Reported by Health Visitor (Public Health Nurse) and as Finally Diagnosed by Physician

		Males			Females		
		Reported Status			Reported Status		
		+	—	Total	+	—	Total
Final diagnosis	+	15.8	20.3	36.1	9.5	7.7	17.2
	—	3.9	60.0	63.9	3.9	78.9	82.8
Total		19.7	80.3	100	13.4	86.6	100
		$\alpha=3.9/63.9=6\%$			$\alpha=3.9/82.8=5\%$		
		$\beta=20.3/36.1=56\%$			$\beta=7.7/17.2=45\%$		
		$P(T+\|S+)=15.8/19.7=0.80$			$P(T+\|S+)=9.5/13.4=0.71$		
		$P(T+\|S-)=20.3/80.3=0.25$			$P(T+\|S-)=7.7/86.6=0.09$		

always less than unity in the presence of misclassification.

Finally, some comment should be made on the assumption that the same values of α and β apply to both populations. For an unbiased measuring instrument, such as a screening blood-sugar test for diabetes,[6] the use of a constant screening level should lead to constant probabilities α and β of misclassifying true negative and positive diabetics, in various populations. In other situations, biased measuring instruments may be used, which give different values of α and β in different populations. If this situation is anticipated, Diamond and Lilienfeld quite rightly suggest that a study of misclassification should be built into the survey. There is no theoretical reason to suppose, however, that the biased values will be such as to make $P(T+|S+)$ and $P(T+|S-)$ constant. For example, in a survey of chronic bronchitis among the citizens of Newcastle upon Tyne,[3] the initial screening for symptoms was carried out by health visitors (public health nurses) in the homes of some 4,000 sample adults. Some bias as between male and female respondents might be expected. This bias appears in Table 3 as the difference between the values of β: 56 per cent for men and 45 per cent for women. Men with bronchitis were somewhat more reluctant than women to admit to their symptoms to the female investigator. It is clear, however, that in the presence of this bias, the alternative probabilities suggested by Diamond and Lilienfeld differed by much more than this. $P(T+|S-)$ was 0.25 for men, but only 0.09 for women.

Acknowledgments—I have discussed the ideas in this paper with several people in the past, including Professor B. G. Greenberg, Chapel Hill, N. C., who has taught the same method as this for a number of years.

REFERENCES

1. Diamond, E. L., and Lilienfeld, A. M. Effects of Errors in Classification and Diagnosis in Various Types of Epidemiological Studies. A.J.P.H. 52:1137–1144, 1962.
2. Bross, I. Misclassification in 2 x 2 Tables. Biometrics 10:478–486, 1954.
3. Ogilvie, A. G., and Newell, D. J. Chronic Bronchitis in Newcastle upon Tyne. Edinburgh and London, England: Livingstone, 1957.
4. Lilienfeld, A. M., and Graham, S. Validity of Determining Circumcision Status by Questionnaire as Related to Epidemiological Studies of Cancer of the Cervix. J. Nat. Cancer Inst. 21:713–720, 1958.
5. Wynder, E. L.; Cornfield, J.; Shroff, P. D.; and Doraiswami, K. R. A Study of Environmental Factors in Carcinoma of the Cervix. Am. J. Obst. & Gynec. 68:1016–1052, 1954.
6. Blumberg, Mark S. Evaluating Health Screening Procedures. Operations Res. 5:351–360, 1957.

Dr. Newell is senior lecturer in biostatistics, University of Durham, the Medical School, King's College, Newcastle upon Tyne, England.